

Synaptic Four

White Paper

März 2026

Ferrum: GA4GH-konforme Infrastruktur nach dem Prinzip „Erst verstehen, dann bauen“

Wie Synaptic Four eine Lücke im GA4GH-Ökosystem identifiziert und das fehlende Werkzeug gebaut hat — an der Schnittstelle von Präzision, offenen Standards und KI-gestützter Softwareentwicklung.

Autorin: Synaptic Four Technical Team · Stuttgart, Deutschland · synapticfour.com

Zusammenfassung

Ferrum ist eine vollständige Implementierung des GA4GH-Standardstacks — TRS, DRS, WES, TES, htsgget, Beacon, Passports und Crypt4GH — entwickelt von Synaptic Four als erstes eigenes Infrastrukturprojekt. Dieses White Paper beschreibt die Motivation, die Architektur, den Deployment-Ansatz und die gewonnenen Erkenntnisse. Ferrum entstand nicht auf Auftrag, sondern weil wir nach einem solchen Werkzeug gesucht, es nicht gefunden und es schließlich selbst gebaut haben. Wir beschreiben außerdem HelixTest, eine unabhängig entwickelte Open-Source-Konformitätssuite für GA4GH, die bei jedem Commit gegen Ferrum läuft. Dieses Dokument dient zwei Zwecken: als technische Übersicht für potenzielle Partner und als konkretes Beispiel dafür, wie Synaptic Four an komplexe Infrastrukturarbeit herangeht.

1. Ursprung und Motivation

1.1 Eine Lücke im Ökosystem

Die GA4GH hat über Jahre hinweg eine kompositionsfähige Reihe von APIs für interoperable genomische Daten und Berechnungen definiert. Die Spezifikationen sind durchdacht und weit verbreitet referenziert. In der Praxis fehlt jedoch etwas Wesentliches: eine einzige, deploybare Laufzeitumgebung, die diese Standards zu einem kohärenten Ganzen zusammenführt.

Als Synaptic Four begann, in diesem Bereich zu arbeiten, suchten wir nach vorhandenen Werkzeugen. Was wir fanden, war Fragmentierung: TES-Implementierungen ohne DRS-Integration, WES-Systeme ohne TRS-Anbindung, Crypt4GH-Deployments ohne Pipeline-Benchmark und Beacon-Knoten ohne Ausführungsschicht. Jede Komponente war isoliert gelöst. Das verbundene System — eines, das von der Workflow-Entdeckung über verschlüsselten Datenzugriff bis zur reproduzierbaren Ausführung in einem einzigen Stack führt — existierte in deploybarer Form nicht.

Die GA4GH hat diese Lücke selbst anerkannt. Die strategische Road Map benennt den Bedarf an produktionsreifen Implementierungen. Das TES-API-Paper (Ellrott et al., 2024) stellt fest, dass der Bedarf an einer Low-Level-API zur Verbindung von Task-Execution und Workflow-Engines als Blocker für Cloud-Interoperabilität identifiziert wurde. Die Lücke ist real und dokumentiert. Ferrum ist unsere Antwort darauf.

„Wir haben nach diesem Werkzeug gesucht. Wir haben es nicht gefunden. Also haben wir es gebaut.“

1.2 Die bestehende Werkzeuglandschaft

Mehrere GA4GH-Komponentenimplementierungen existieren und werden aktiv gepflegt. TESK (EMBL-EBI / ELIXIR) und Funnel (OHSU) implementieren TES. cwl-WES (ELIXIR Cloud & AAI) implementiert WES für CWL-Workflows gegen TES-Backends. WESkit (DKFZ / Sanger Institute) implementiert WES für Nextflow und Snakemake in HPC-Umgebungen. Das GA4GH Starter Kit bietet Java-Referenzimplementierungen von DRS und WES. Galaxy integriert TRS, DRS und TES auf seiner Plattform.

Diese Werkzeuge sind wertvoll. Was sie gemeinsam haben: Sie sind einzelne Komponenten oder Spezialintegrationen. Keines von ihnen kombiniert TRS, DRS, WES, TES, htsgat, Beacon, Passports und Crypt4GH in einer einzigen Laufzeitumgebung mit einheitlichem Gateway und gemeinsamer Authentifizierung. Keines bietet eine Cross-Service-Konformitätssuite in CI. Und keines ermöglicht eine One-Command-Deployment-Strecke, die den gesamten Stack End-to-End ausführt.

Ferrum unterscheidet sich im Umfang: Es ist eine Plattform, kein Einzelbaustein. Einzelne Dienste können selektiv aktiviert werden für Mix-and-Match-Deployments in bestehender Infrastruktur. Aber alle leben in einer Codebasis, teilen ein Gateway und werden gemeinsam validiert.

2. Was Ferrum ist

2.1 Systemarchitektur

Ferrum implementiert alle sechs großen GA4GH-API-Komponenten — plus Passports und Crypt4GH — in einer einzigen, in Rust geschriebenen Laufzeitumgebung. Ein einheitliches Gateway übernimmt Routing, Authentifizierung und TLS. Alle Dienste teilen einen PostgreSQL-Metadaten Speicher und ein konfigurierbares Objekt-Storage-Backend.

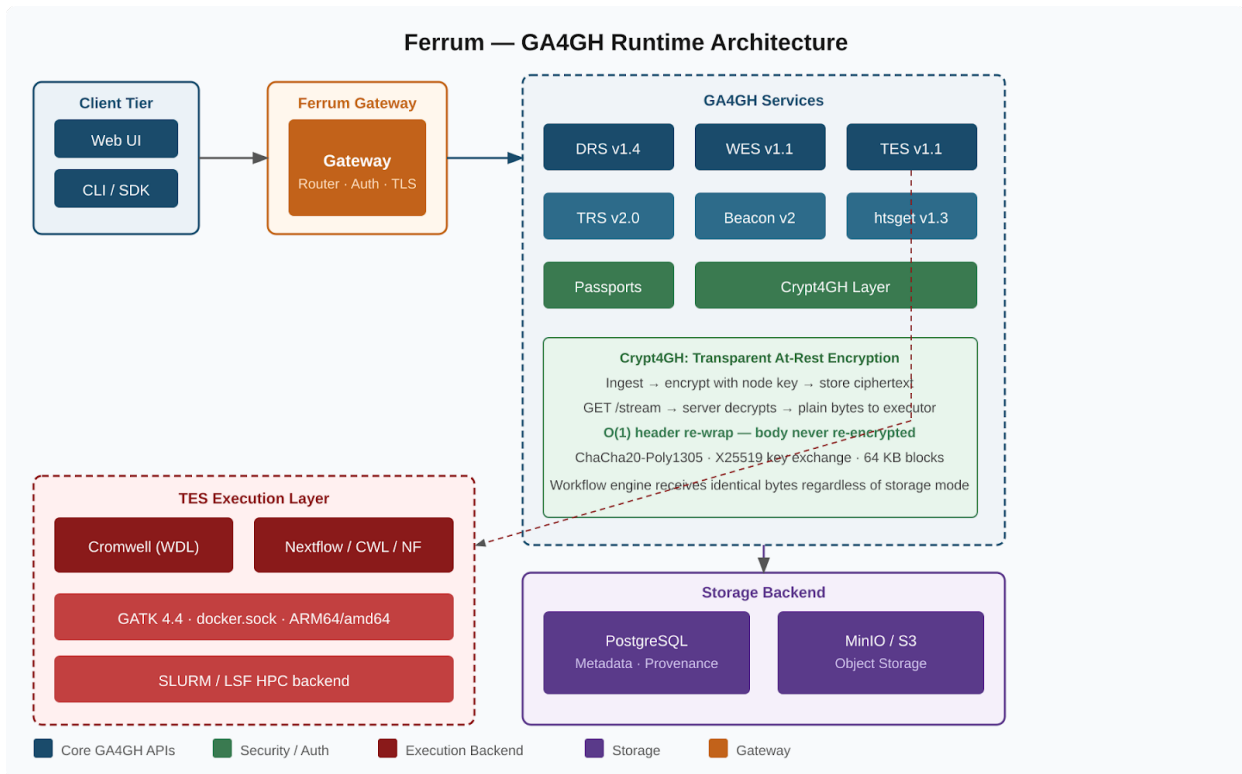


Abbildung 1. Ferrum-Gesamtarchitektur. Alle GA4GH-API-Komponenten sind unter einem einzigen Rust-Gateway integriert. Die Crypt4GH-Schicht bietet transparente Verschlüsselung im Ruhezustand auf DRS-Ebene. Das Ausführungsbackend unterstützt Cromwell, Nextflow, CWL, Snakemake, SLURM und LSF.

2.2 Storage-Backends

Ferrum unterstützt mehrere Storage-Backends. Die Benchmark-Läufe in diesem Paper verwenden eine lokale MiniIO-S3-kompatible Instanz. Weitere unterstützte Optionen:

- POSIX-Dateisystem — für On-Premises-HPC ohne S3
- Beliebiger S3-kompatibler Anbieter — AWS, Hetzner, Cloudflare R2 und andere
- OpenDAL — für Azure Blob, GCS und weitere Backends

Das Storage-Backend ist von der GA4GH-API-Schicht entkoppelt. Workflow-Executoren interagieren ausschließlich mit DRS-Endpunkten; der darunterliegende Speicher ist für sie unsichtbar.

2.3 Selektives Deployment und Ferrum Lab Kit

Ferrum ist kein Alles-oder-Nichts-System. Jeder GA4GH-Dienst lässt sich einzeln aktivieren oder deaktivieren. Eine Institution, die nur Beacon v2 für die Kohorten-Discovery benötigt, kann nur Beacon deployen. Ein Labor, das DRS und WES/TES, aber kein htsget braucht, lässt es einfach weg. Die Dienste teilen Gateway und Auth-Schicht bei gemeinsamer Nutzung, sind aber auch eigenständig nutzbar.

Für kleinere Labore und ELIXIR-Node-Kandidaten pflegt Synaptic Four das Ferrum Lab Kit — eine separate Deployment- und Integrationsschicht, die einen strukturierten Einstieg ermöglicht, ohne die gesamte Plattform benötigt. Lab Kit richtet sich an universitäre Forschungsgruppen, GDI-Nationalknotenpartizipanten, GHGA-Dateneinreichende und NFDI-Verbundprojekte, die GA4GH-konforme Dienste auf SLURM oder einem einzelnen

Server benötigen. Es integriert ELIXIR LS Login out-of-the-box und generiert Docker-Compose-, Helm- oder systemd-Konfigurationen aus einer einzigen TOML-Datei.

Ferrum Lab Kit: github.com/SynapticFour/Ferrum-Lab-Kit · Einstiegspunkt für Labore und ELIXIR-Node-Kandidaten · BUSL-1.1

2.4 Crypt4GH: Transparente Verschlüsselung

Daten werden bei der Aufnahme verschlüsselt gespeichert und beim Abruf via GET /stream vom Gateway transparent entschlüsselt. Für den Workflow-Executor sind verschlüsselte und unverschlüsselte Objekte nicht unterscheidbar — die Verschlüsselung ist vollständig transparent. Das Header-Re-Wrapping für empfangerspezifische Schlüssel erfolgt in O(1): Nur der Crypt4GH-Header wird neu verpackt, unabhängig von der Dateigröße. Eine 500-GB-BAM-Datei und eine 1-KB-Referenzdatei werden in derselben Zeit re-wrapped.

2.5 Multi-Engine-Ausführung

Ferrum unterstützt Cromwell (WDL), Nextflow, CWL und Snakemake als WES-Backends, geroutet über TES. HPC-Scheduling via SLURM und LSF wird ebenfalls unterstützt. Dieselben Workflows, Daten und Benchmark-Infrastruktur laufen auf allen Engines. Die Engine-Portabilität wird in CI durch HelixTest getestet und in den nachfolgenden Benchmark-Ergebnissen demonstriert.

3. Validierung

3.1 Ausführungsfluss

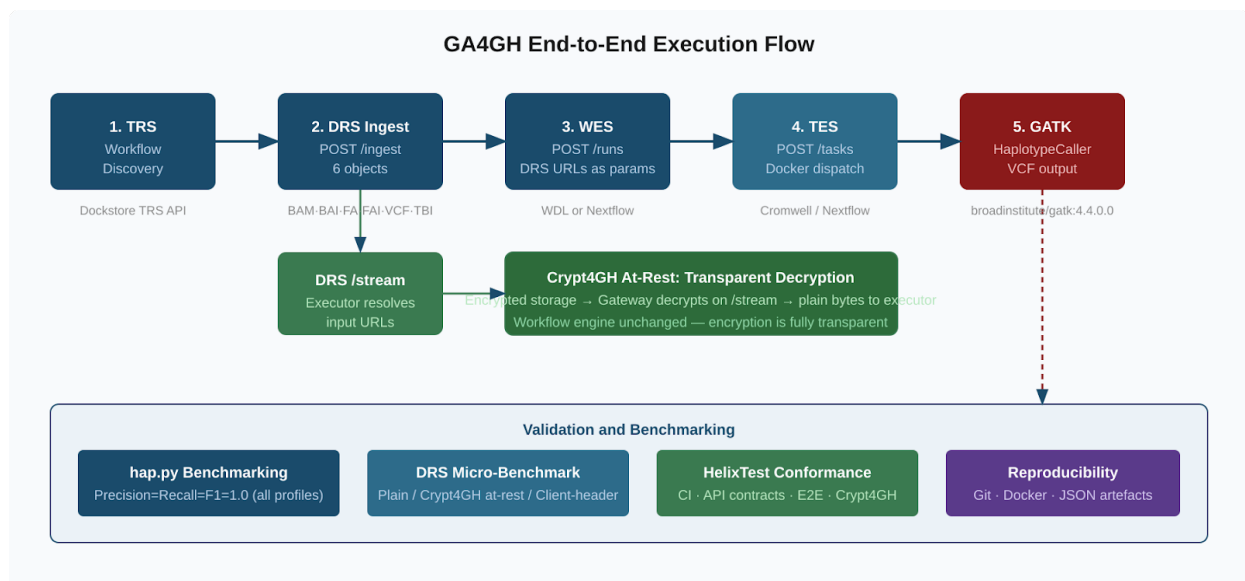


Abbildung 2. GA4GH-End-to-End-Ausführungsfluss: TRS-Abruf → DRS-Ingest → WES-Übermittlung → TES-Dispatch → GATK-Ausführung → hap.py-Validierung. Die Crypt4GH-Entschlüsselung erfolgt transparent auf DRS-/stream-Ebene.

3.2 Benchmark-Ergebnisse

Wir haben einen GATK-HaplotypeCaller-Workflow auf einem synthetischen GIAB-Datensatz ausgeführt und mit hap.py bewertet. Alle Profile — plain, Crypt4GH at-rest, Cromwell und Nextflow — erzielen Precision=Recall=F1=1,0. Das Crypt4GH-Profil erzeugt ca. 10 % mehr Gesamtlaufzeit, was auf die Verschlüsselung beim Ingest zurückzuführen ist, nicht auf das Streaming.

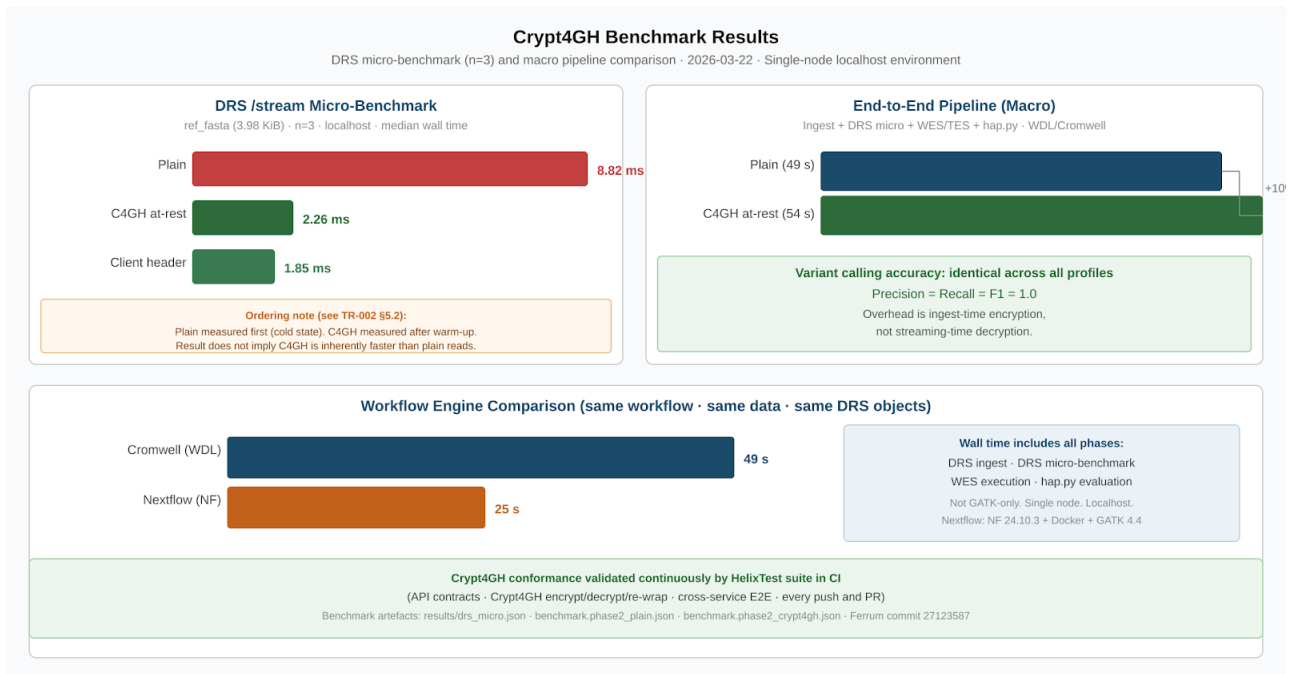


Abbildung 3. Benchmark-Ergebnisse. DRS-Mikro-Benchmark (oben links): drei Zugriffsmodi, n=3 Wiederholungen. End-to-End-Makro-Pipeline (oben rechts): plain vs. Crypt4GH at-rest. Engine-Vergleich (unten): Cromwell vs. Nextflow auf identischem Workflow und Datensatz.

Profil	Engine	Laufzeit (s)	Precision	Recall	F1
Plain Ingest	Cromwell/WDL	49	1,0	1,0	1,0
Crypt4GH at-rest	Cromwell/WDL	54	1,0	1,0	1,0
Plain Ingest	Nextflow	25	1,0	1,0	1,0

4. HelixTest: Konformität als Kernprinzip

Ferrum wird bei jedem Commit durch HelixTest validiert — eine unabhängig entwickelte, Apache-2.0-lizenzierte, in Rust geschriebene GA4GH-Konformitätssuite. Sie deckt API-Verträge, Workflow-End-to-End, Cross-Service-Integration (TRS → DRS → WES → TES → Beacon), GA4GH Passports/OIDC-Auth, htsgit-Streaming und Crypt4GH-Kryptografiekorrektheit ab.

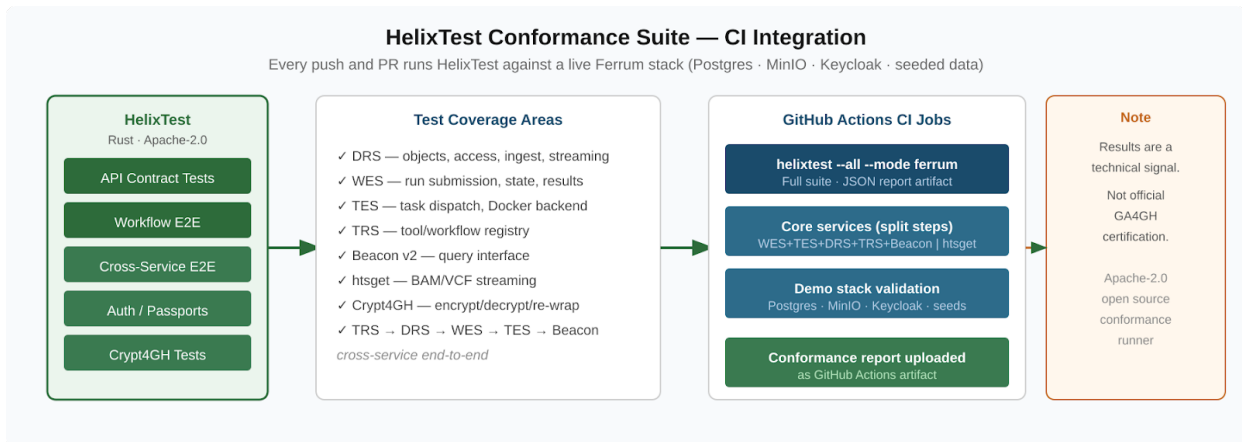


Abbildung 4. HelixTest-CI-Pipeline. Jeder Push löst zwei Jobs aus: vollständige Suite und getrennte Kerndienstvalidierung. Alle Ergebnisse werden als GitHub-Actions-Artefakte hochgeladen.

HelixTest ist ein eigenständiges Werkzeug, das von jeder GA4GH-kompatiblen Plattform genutzt werden kann. Es ist ein Beitrag zum breiteren Ökosystem: ein wiederverwendbarer, CI-freundlicher Konformitäts-Runner für alle sechs GA4GH-API-Bereiche. Bestehende Implementierungen wie cwl-WES, WESkit, Funnel oder TESK verfügen nicht über eine äquivalente Cross-Service-Testsuite, die gegen einen Live-Stack in CI läuft.

5. Wie Synaptic Four arbeitet

5.1 KI als Werkzeug, nicht als Abkürzung

Synaptic Four setzt KI-gestützte Entwicklung als zentralen Bestandteil seiner Entwicklungspraxis ein. Wir sind darüber transparent, weil Methodentransparenz eine berufliche Verpflichtung ist, keine Schwachstelle. KI-Werkzeuge beschleunigen das Gerüst standardskonformer Komponenten und unterstützen die Codegenerierung für repetitive, aber präzise Aufgaben — wie die Implementierung von GA4GH-API-Endpunkten, die exakt einem Schema entsprechen müssen.

Was das nicht bedeutet: KI trifft Architekturentscheidungen. Jede Integrationsstelle, jede Designentscheidung, jeder Validierungsschritt in Ferrum wurde explizit durchdacht und überprüft. Wir nutzen KI als fähigen Kollaborateur, der Code schreibt; das ingenieurmäßige Urteilsvermögen und die Verantwortung liegen bei uns.

„KI ist ein Werkzeug. Die Verantwortung für Korrektheit, Design und Konsequenzen liegt beim Ingenieur. Diese Linie halten wir klar.“

5.2 Präzision und Transparenz

Unsere Benchmark-Ergebnisse sind ehrlich in Bezug auf den Umfang. Der Validierungsdatensatz ist klein und synthetisch. Wir extrapolieren nicht über das Gemessene hinaus. Einschränkungen sind in den technischen Begleitberichten explizit dokumentiert. Wir glauben, dass dies die einzige glaubwürdige Arbeitsweise in einem Bereich ist, in dem Reproduzierbarkeit ein Kernwert ist.

5.3 Neurodiversität als Stärke

Synaptic Four wurde von Menschen im Autismus-Spektrum gegründet — nicht als Initiative von außen, sondern aus eigener Erfahrung. Präzision, strukturiertes Problemlösen und Detailgenauigkeit sind keine Marketingpositionierung, sondern beschreiben, wie die Gründer:innen natürlicherweise arbeiten. Diese Eigenschaften passen gut zu anspruchsvoller Bioinformatik-Infrastruktur, wo Ungenauigkeiten und Abkürzungen falsche Ergebnisse in folgenreichen Kontexten produzieren.

Langfristig möchten wir mit Entwickler:innen in Ghana zusammenarbeiten: qualitativ hochwertige, kosteneffiziente IT-Lösungen für Kunden in Europa und den USA, verbunden mit Ausbildungs- und Beschäftigungsmöglichkeiten vor Ort. Technische Exzellenz und soziale Verantwortung schließen sich für uns nicht aus.

6. Für potenzielle Partner

Ferrum und Ferrum Lab Kit sind heute für nicht-kommerzielle Forschungsnutzung (BUSL-1.1) deploybar. Organisationen, die GA4GH-konforme Infrastruktur aufbauen — Forschungsinstitutionen, ELIXIR-Node-Kandidaten, nationale Genomikprogramme, klinische Datenplattformen — können mit Synaptic Four in folgenden Bereichen zusammenarbeiten:

- Deployment und Integration von Ferrum oder ausgewählten Komponenten
- GA4GH-Compliance-Beratung und Infrastrukturdesign
- Bioinformatik-Softwareentwicklung
- Training zu GA4GH-Standards, Interoperabilität und sicherem genomischen Datenzugang

Kontakt: contact@synapticfour.com · synapticfour.com · Stuttgart, Deutschland

Synaptic Four · Stuttgart, Deutschland · contact@synapticfour.com · synapticfour.com

Dieses White Paper darf für nicht-kommerzielle Zwecke mit Quellenangabe frei geteilt werden.